14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

TASK IMPACT ON THE VISUAL ATTENTION IN SUBJECTIVE IMAGE QUALITY ASSESSMENT

A. NINASSI[1,2], O. LE MEUR[1], P. LE CALLET[2], D. BARBA[2], A. TIREL[2]

[1] THOMSON R&D
1 Avenue Belle Fontaine
35511 Cesson-Sevigne, France

[2] IRCCyN UMR 6597 CNRS
Ecole Polytechnique de l'Universite de Nantes
rue Christian Pauc, La Chantrerie, 44306 Nantes, France

ABSTRACT

Visual attention is a main feature of the Human Visual System (HVS). Knowing and using the mechanisms of the visual attention could help to the improvement of image quality assessment methods. But, which kind of saliency should be taken into account? A free-task visual selective attention or a quality oriented visual selective attention. We recorded and evaluated the discrepancy between these two types of oculomotor behavior. The results show the impact of the viewing task on visual strategy.

1. INTRODUCTION

A recurrent issue of the imaging industry is to be able to control the visual quality of their products. The only way to cope with this problem is to have an accurate quality metric. It must be an automatic metric which provides computed quality scores well correlated with the ones given by human observers. Image quality assessment has been extensively studied these past few decades. The most efficient metrics are based on the HVS [1] [4] [3] . One of the most important component is called the visual attention. It is divided into the top-down and the bottom-up processes. The bottom-up process is guided by the low-level features of the viewed stimuli, and the top-down process is guided by high level cognitive factors [2].

Studying the visual attention is thus a way to improve image quality assessment. For example, an artifact that appears on a region of interest is much more annoying than a degradation appearing on inconspicuous area.

In this paper, we attempt to answer the following question: what are the visual strategy differences when a free viewing task or a quality-task are considered? In order to deal with this issue, eye tracking experiments are conducted both in free-task and in quality-task. This paper is decomposed as follows. Section 2 is devoted to the eye tracking experiments description. The results are analyzed in the section 3. Finally, results are summarized and some conclusions are drawn.

2. EYE TRACKING EXPERIMENTS

2.1 Eye tracking apparatus

In order to track and record real observers eye movements, experiments have been performed with a dual-Purkinje eye tracker from *Cambridge Research Corporation*. The eye tracker is mounted on a rigid EyeLock headrest that incorporates an infrared camera, an infrared mirror and two infrared illumination sources. To obtain accurate data regarding the diameter of the subjects's pupil a calibration procedure is needed. The calibration requires the subject to view a number of screen targets from a known distance. Once the calibration procedure is complete and a stimulus has been loaded, the system is able to track a subject's eye movement. The camera recorded a close-up image of the eye. Video was processed in real-time to extract the spatial location of the eye position. Both Purkinje reflections are used to calculate the location. The guaranteed sampling frequency is 50 Hz and the accuracy is about 0.5 degree.

2.2 Subjects

Twenty unpaid subjects participated to the experiments. They came from the university of Nantes. All had normal or corrected to normal vision. All were inexperienced observers (not expert in video processing) and naive to the experiment. Before each trial, the subject's head was positioned so that their chin rested on the chin-rest and their forehead rested against the head-strap. The height of the chin-rest and head-strap was adjusted so that the subject was comfortable and their eye level with the center of the presentation display.

2.3 Free viewing task

Twenty pictures of various contents have been selected. Ten pictures present numerous artifacts. Each picture was presented to subjects in a free-viewing task during 8s. A gray picture is displayed during 3s between two test pictures. Each trial began with the calibration of the eye tracker. Experiments were conducted in normalized conditions (ITU-R BT 500-10). Image resolution was 512×512. They are displayed at viewing distance of four times the height of the picture (80 cm).

Subjects were instructed to "look around the image". The free viewing condition is mandatory to lessen the top-down effects. It is required that visual attention was mainly driven by the low level visual features.

2.4 Picture quality assessment task

In this eye tracking experiment, participants have to assess the quality of a picture. The fact that a particular task is assigned will likely alter the oculomotor behavior.

To perform the picture quality evaluation, the standardized method DSIS (Double Stimulus Impairment Scale) is used. In DSIS, each observer views an unimpaired reference picture followed by its impaired version. Observer then rates the impaired pictures using a scale containing 5 scores:

- imperceptible: impairments are imperceptible;
- not annoying: observers are not annoyed by the impairments;
- same: there is no perceptible difference between the two pictures;

- annoying: observers are annoyed by the impairments;
- very annoying: observers are very annoyed by the impairments.

To avoid lots of eye tracker calibration, the observers do not give their quality scores with a device such as keyboard or voice, but with their eyes. The observer just gaze the scoring screen area (figure 1) corresponding to his choice. The chosen area becomes darker, and then the observer has to validate or to correct his choice by directing his gaze to the corresponding screen area.

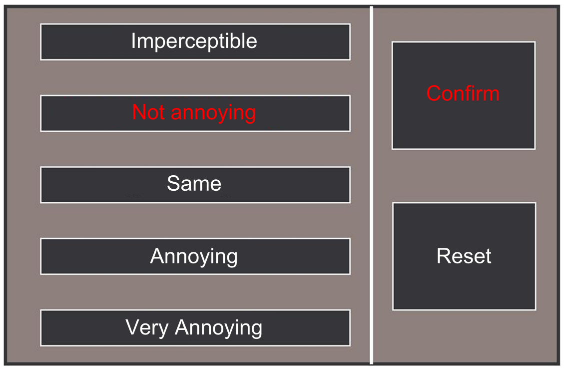

Figure 1: Scoring screen displays between two assessments.

Ten unimpaired pictures are used in this experiments. These pictures are impaired by a JPEG, JPEG2000 compression or by a blurring filter. 120 impaired pictures are obtained.

2.5 A human fixation density

From the collected data, a fixation map is computed for each observer and for each picture. It encodes the saliency degree of each spatial location of the picture. This kind of map is often compared to a landscape map [6] consisting of peaks and valleys. A peak, indicating the number of fixations, represents the observer's regions of interest. A saliency map SM^k for an observer k is given by

$$SM^k(x,y) = \sum_{j=1}^{NbData} \Delta(x - x_j, y - y_j), \qquad (1)$$

$NbData$ is the number of data collected by the eyetracker, and Δ is the Kronecker delta.
To determine the most visually important regions, all the fixation maps are merged yielding to an average fixation map SM. The average saliency map is given by

$$SM(x,y) = \frac{1}{N} \sum_{k=1}^{N} SM^k(x,y), \qquad (2)$$

N is the number of observers. The average saliency map encodes the most attractive part of a picture when a large panel of observers is considered. Finally, the average saliency map is smoothed with a 2D Gaussian filter, given a density saliency sequence DM:

$$DM(x,y) = SM(x,y) * g_\sigma(x,y), \qquad (3)$$

The standard deviation σ is determined in accordance with the accuracy of the eye-tracking apparatus. This filtering also deals with the fact observers gaze at a particular areas rather than at a precise point.

3. QUANTITATIVE ANALYSIS

3.1 Duration of fixation according to the task

The average fixation duration is computed when the following cases are considered:

1. the original picture is viewed by observers in a free-task configuration,
2. the original picture is viewed by observers in a quality-task configuration. This picture is just displayed before the impaired picture,
3. the impaired picture is viewed by observers in a quality-task configuration.

From the collected data, an average fixation duration is computed for each observer and for each picture. Two fixation phases are temporally separated by a saccade phase. To obtain the average fixation duration for a picture, we work out the average of the average fixation durations per observer for this picture.

Figure 2 gives the average fixation duration in the three aforementioned cases. This analysis indicates that the average fixation durations are similar when the free-task and the quality-task configuration (with the impaired picture) are considered. In this case, the oculomotor behavior is not disturb by the task. It is important to stress that this result does not mean that observers pay attention to the same locations. It only means that one parameter of the visual strategy is unmodified.
Considering the quality-task configuration with the original picture, the duration fixations are significantly longer than the previous ones. In this case, the oculomotor behavior is clearly modified. A possible explanation lies in the fact observers endeavor to accurately memorize some parts of the picture. Spatial memory seems here important to achieve the proposed task.

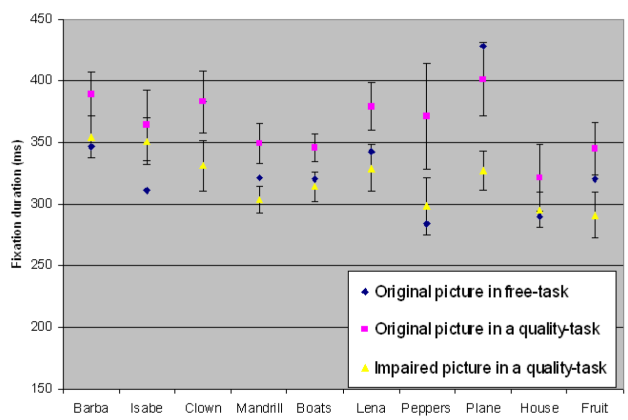

Figure 2: Fixation durations assessment for the free-task and the quality-task (The viewing duration is 8s). A 95% confidence interval is given for each case.

3.2 Correspondence between the different saliency maps

3.2.1 Metrics

To test the correspondence between the different saliency density maps, two metrics are used: the Kullback-Leibler divergence and the correlation coefficient. The former assess the degree of dissimilarity that potentially exists between two probability density functions. The Kullback-Leibler divergence, noted $KL()$ is given below:

$$KL(p|h) = \sum_x p(x) Log(\frac{p(x)}{h(x)}) \quad (4)$$

with,
h and p are the probability density functions. When the two probability densities are strictly equal, the KL value is zero.
The latter, the linear correlation coefficient, noted CC, measures the strength of a linear relationship between two variables. It has a value between -1 and $+1$. When the correlation is close to $+/-1$, there is an almost perfectly linear relationship between the two variables. The correlation coefficient CC is given by

$$CC(p,h) = \frac{cov(p,h)}{\sigma_p \sigma_h} \quad (5)$$

p and h represent the saliency density maps
$cov(p,h)$ is the covariance value between p and h.
σ_p and σ_h represent the standard deviation for the saliency density maps, p and h, respectively.

3.2.2 Average observer behavior

Figure 3 illustrates the four measures we have done:

- test (A), *Reference in quality-task versus reference in free-task*: in this first test, we focus on the influence of the task on the oculomotor behavior [7]. Do the observers look at the same area?
- test (B), *Reference in quality-task versus first reference in quality-task*: the objective here is to show (or not) that observers adapt theirs visual strategies to inspect the original picture in a quality-task. Do they learn something in order to refine their quality judgment?
- test (C), *Degraded quality task versus reference free task*: it is well known that the task acts on the allocation of attention. But we do not know to what extent a task modify the visual attention. This issue is here tackled by comparing saliency maps coming from a free-task and from a quality-task. Moreover, do the artifacts modify the saliency maps?
- test (D), *Degraded quality task versus its associated reference quality task*: in a DSIS method, is the visual strategy the same for the reference and the impaired pictures?

Results of the two first analysis are displayed on figure 4 whereas the two last analysis are given by figure 6.

As expected, the degree of dissimilarity between two saliency maps is important when two different tasks are considered (see figure 4). The KL value is in the range $[0.3, 0.5]$. The same trend is observed for the CC value that is in the range $[0.77, 0.92]$. To go deeper in this analysis, the difference between the two saliency maps is computed. Several examples are given in figure 5. Yellow areas are seen with the same intensity in both the free-task and the quality-task.

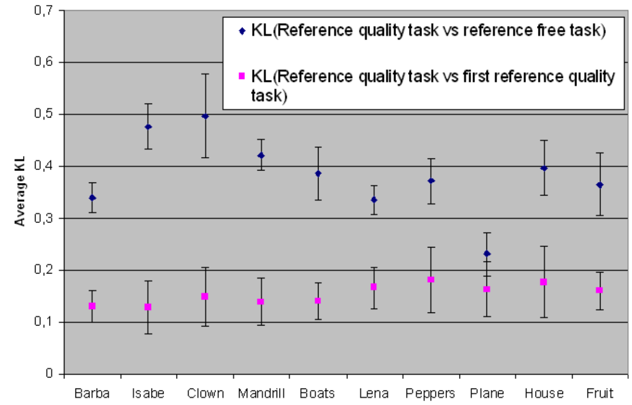

Figure 4: Average Kullback-Leibler divergence computed for each original picture. As shown in figure 3, the KL value is computed on one hand between the density map of the original picture in a quality-task and the density map of the original picture in a free-task, and on the other hand between the density map of original picture in a quality-task and the first density map coming from the first original picture viewed in quality-task.

Red areas correspond to the areas that are more inspected in quality-task.
For the picture Clown and for a quality-task, the most important differences concern clown's face, head and his hand. Observers take more time to see these areas (this is coherent with the result of figure 2). Concerning the picture Boats, the error map indicates that the main difference is located on the boat's name.

Second result of figure 4 concerns the adaptation of the visual strategy for a quality task. As observers saw several times the same unimpaired picture, the short term memory and the observers capacities to learn how assessed the picture quality (for example, to assess the picture quality, it is preferable to scan flat areas rather than textured areas) can likely modified the visual strategy. Although that it was reasonable to think that observers become more and more competitive, the results indicate that this hypothesis is wrong. Both the degree of dissimilarity and the confidence interval are weak. The KL and the CC values are respectively in the range $[0.12, 0.18]$ and $[0.9, 0.97]$.

Figure 6 allows to tackle two points: what are the differences between the free-task and the quality-task when the impaired picture is considered. The second points refers to the similarity of the visual strategy when an unimpaired and impaired pictures are considered. In others words, does an artifact have the capacity to attract or to significantly modify the visual attention?
Concerning the first point, results indicates that there exists a significant difference between the visual strategy that is deployed for a free-task and a quality-task (the results of figure 4 are retrieved). The KL and the CC values are respectively in the range $[0.42, 0.95]$ and $[0.66, 0.9]$. Moreover, the confidence intervals are important compared to those of figure 4. It means that the type of degradation (Blur, JPEG, JPEG2000) has a significant influence. Figure 7 gives three difference maps (difference between the saliency map obtained in free-task and the saliency map obtained in quality-task). It is noticeable that there is no striking

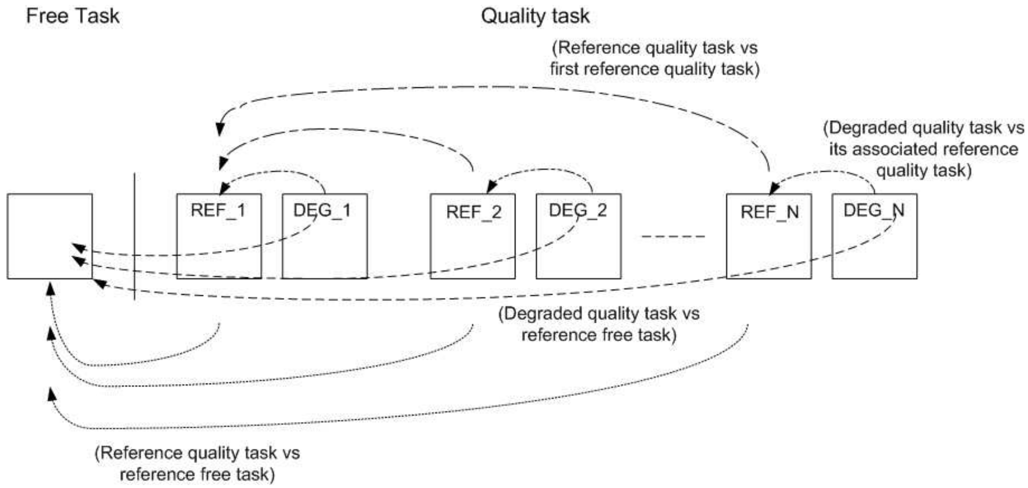

Figure 3: This schema summarizes the quantitative analysis that have been performed.

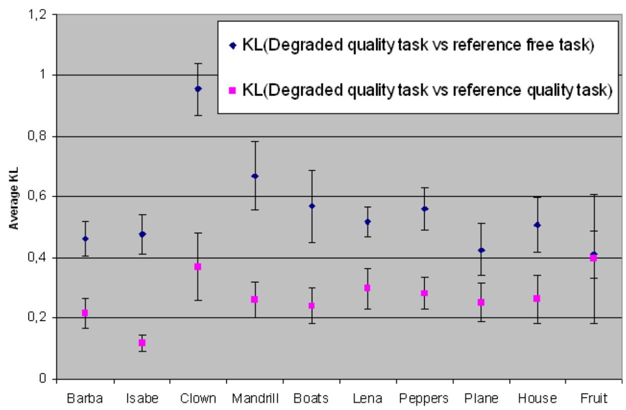

Figure 6: Average Kullback-Leibler divergence computed for each original picture, whatever the degradation is. The KL value is computed in on hand between the density map of the degraded picture in a quality-task and the density map of the original picture in a free-task, and in the other hand between the density map of degraded picture in a quality-task and the density map coming from its associated reference picture viewed in quality-task.

resemblance between the maps.

The last study concerns the comparison of the visual strategy deployed on the unimpaired and the impaired pictures. The KL and CC values are respectively in the range $[0.11, 0.28]$ and $[0.8, 0.96]$, leading to the conclusion that there is few differences between the two saliency map (stemming from both the unimpaired and the impaired pictures).

4. CONCLUSION

As we expected, a quality task has a significant effect on eye movements. The first result shows that the fixation duration increased on the unimpaired picture used in a quality-task. It means that observers attempt to memorize some parts of the picture. The second important result concerns the variation of the visual strategy throughout the quality test. We show that observers are not more competitive at the end of test than at the beginning. In other word, there is no visual adaptation or task learning. Finally, when a cross comparison is done between the KL values of test (A) (reference quality task versus reference free task) and test (C) (degraded quality task versus reference free task), it is interesting to note the type of degradation modifies the visual strategy [5].
The future study has to deal both with the dependence between visual attention and type of degradation and with the analysis of the error maps.

ACKNOWLEDGEMENTS

The authors would like to thank Fabien Beugne and Guillaume Courtin from the university of Nantes for their contributions to the eye tracking experiments.

REFERENCES

[1] S. Daly, "A Visual Model for Optimizing the Design of Image processing Algorithms," *In IEEE International Conference on Image Processing*,pp. 16–20, 1994.

[2] W. James, "The Principles of Psychology," Holt, New York, 1890.

[3] P. Le Callet and D. Barba, "Perceptual color image quality metric using adequate error pooling for coding scheme evaluation," *SPIE Human Vision and Electronic Imaging*,San Jose, 2002.

[4] P.C. Teo and D. J. Heeger, "Perceptual image distortion," *In Proceedings of ICIP (IEEE International Conference on Image Processing)*, vol. 2 ,pp. 982–986, 1994.

[5] T. Vuori and M. Olkkonen, "The effect of image sharpness on quantitative eye movement data and on image quality evaluation while viewing natural images," *Human Vision and Electronic Imaging*, San Jose, 2006.

[6] D. S. Wooding, "Eye movements of large population : II. Deriving regions of interest, coverage, and similarity using fixation maps," *Behavior Research Methods, Instruments and Computers*, vol. 34(4) ,pp. 509–517, 2002.

[7] A. Yarbus, "Eye movements and vision," *L.A. Riggs, Trans., New-york :Plenum Press*, 1967.

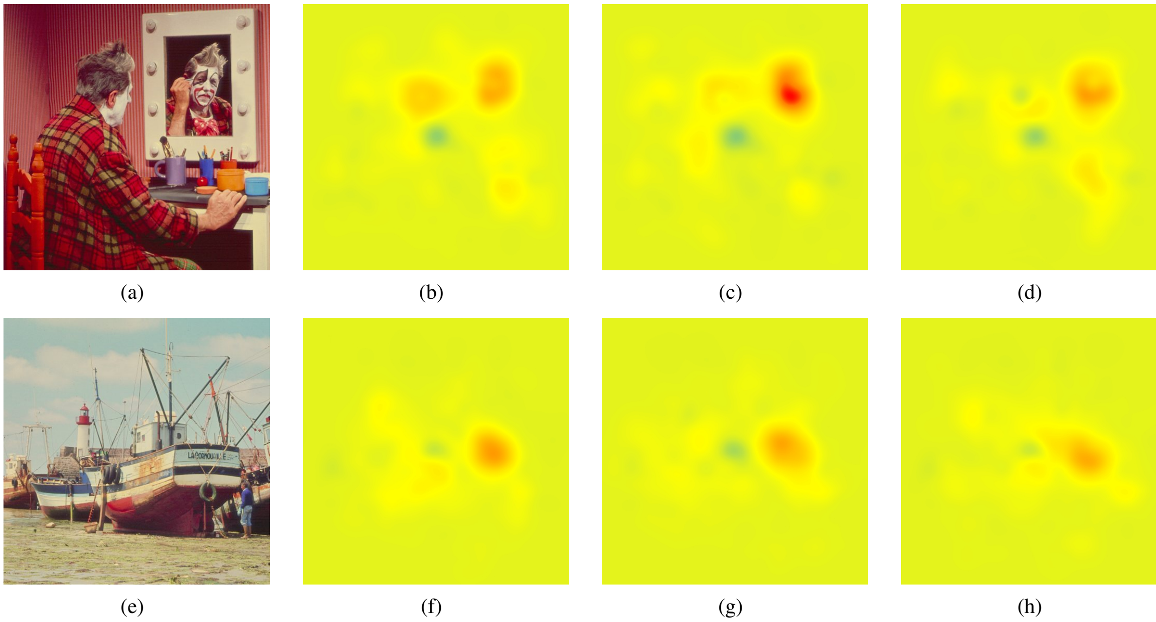

Figure 5: First row:(a) original picture Clown; (b) difference between the free-task saliency map and first reference in quality task; (c) difference between the free-task saliency map and third reference in quality task; (d) difference between the free-task saliency map and fifth reference in quality task. Second row: (e) original picture Boats; (f) difference between the free-task saliency map and first reference in quality task; (g) difference between the free-task saliency map and third reference in quality task; (h) difference between the free-task saliency map and fifth reference in quality task. Red areas are most inspected in quality-task, regarding the free-task. Blue areas are the areas that are neglected in a quality-task.

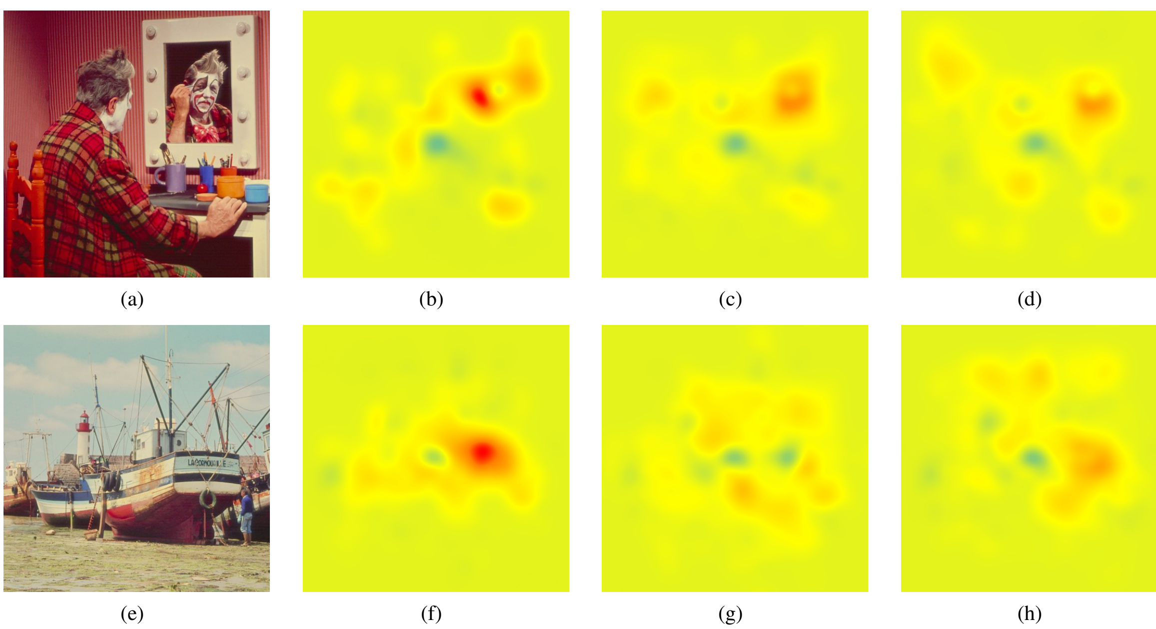

Figure 7: First row:(a) original picture Clown; (b) difference between the free-task saliency map and impaired (JPEG2000) reference in quality task; (c) and (d) difference between the free-task saliency map and two different impaired (JPEG) references in quality task. Second row: (e) original picture Boats; (f) difference between the free-task saliency map and impaired (blurring) reference in quality task; (g) difference between the free-task saliency map and impaired (JPEG2000) reference in quality task; (h) difference between the free-task saliency map and impaired (JPEG) reference in quality task. Red areas are most inspected in quality-task, regarding the free-task. Blue areas are the areas that are neglected in a quality-task.